Managed by Fermi Research Alliance, LLC for the U.S. Department of Energy Office of Science

Space usage monitoring
for distributed heterogeneous data storage systems

Natalia Ratnikova

OPOS seminar

May 3rd, 2016

CMS and WLCG sites

149 storage nodes registered in CMS PhEDEx database:

WLCG dash-board stats:

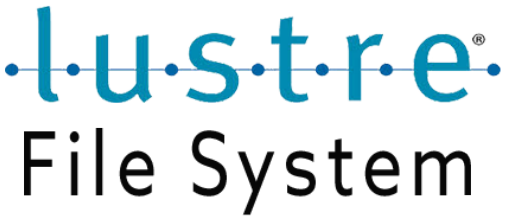
ALICE	ATLAS	CMS	LHCb
137	173	168	85

Fermilab

Storage technologies

StoRM

Storage Resource Manager



DPM

Disk Pool Manager

eos.web.cern.ch

Large Disk Storage at CERN

Data storage at CMS sites

- Total over 100 sites
- Only Tier-1 and Tier-2 sites pledge storage space
- Storage technologies: Castor, dCache, DPM, EOS, Hadoop, LStore, Lustre, StoRM.
- CMS Tier 1 and 2 storage space requirements*

Year	2013	2014	2015	2016
Tier 1 Disk	26,000	26,000	26,000	33,000
Tier 1 Tape	50,000	55,000	74,000	100,000
Tier 2 Disk	26,000	27,000	29,000	38,000

•Increased pileup, higher HLT rate, data parking and scouting
•Volume will grow proportionally to LHC life time
•Phase 2 detector upgrade studies

➡CMS expects severe resource constraints

Evolution of the computing model

- Changed patterns in organized data processing
- Tier 1 disk and tape separation
- AAA xrootd driven data federations
- Dynamic data management
- New data types:
  - MiniAOD
  - phase 2 detector studies
  - parked data
- Diverse user analysis patterns
- Increased share of storage space for users and groups

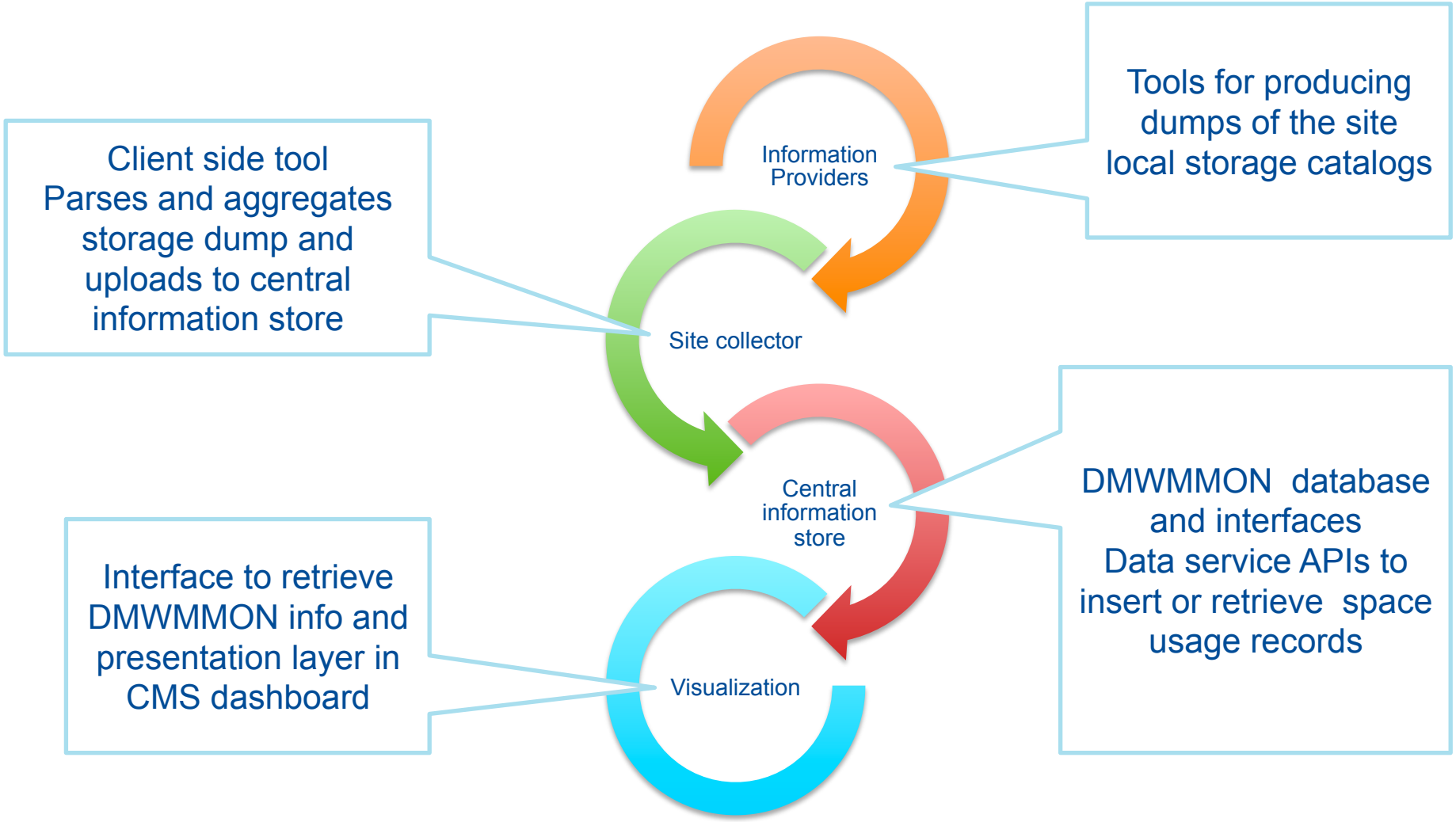
Multiple data placement processes not necessarily aware of each other sharing the same storage resources

Space monitoring for distributed storage

- CMS data live in a **global name space**, addressed by a logical file name (LFN), e.g.:

  /store/data, /store/mc, /store/user, /store/group, …
- Data are accessed by physical file names (PFNs) according to the LFN to PFN translation rules specified in the trivial file catalogs provided by the sites
- Space monitoring allows to track the space occupied by each level under /store across the sites.
- **Main use cases**:
  - Efficient space utilization
  - Fair share between users and groups
  - Resource planning

CMS Space Monitoring system overview



Fermilab

Storage information providers

Storage dumps:

- Storage dumps in SynCat XML format: supported by DPM, dCache
- WLCG recommended format, tools available for EOS, Storm, Lustre…
- Customized formats used at KIT dCache, CERN EOS, FNAL Enstore …
- Storage dumps also used for consistency checks detecting grey data

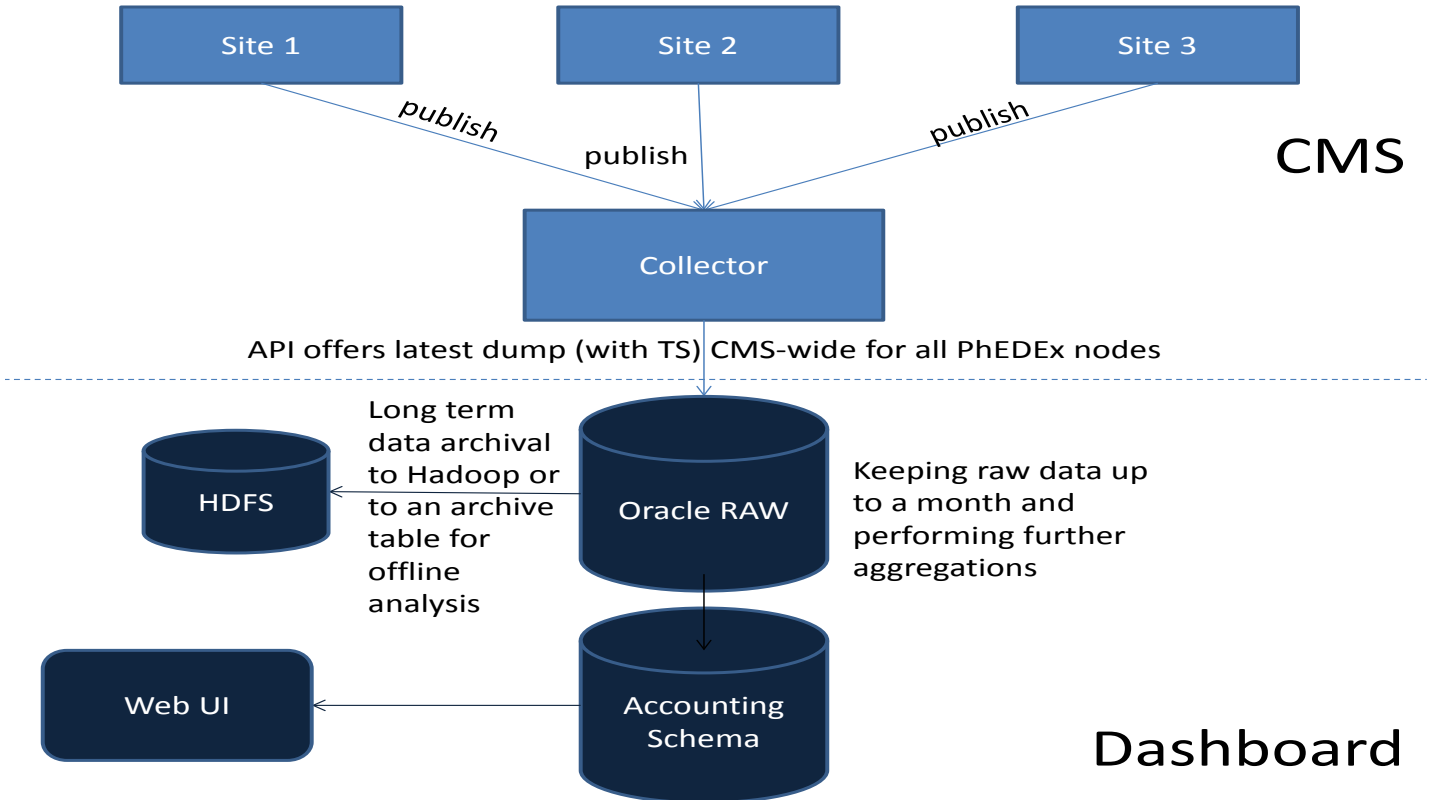
Alternatives: aggregation on DB level, crawling namespace mounted on the grid workers from the grid jobs, space reporting via DAV…

- The storage providers have agreed that the way to provide usage information via HTTP/DAV is via RFC4331.
- NB – this is related to PATHS and **not** to SPACE TOKENS.
- Migration depends upon existing conventions linking these concepts

<d:multistatus xmlns:d="DAV:"> <d:href>/dpm/cern.ch/home/dteam</d:href>
... <d:prop>
<d:quota-available-bytes>282476624607</d:quota- available-bytes>
<d:quota-used-bytes>4212442401</d:quota-used-bytes> </d:prop>
... </d:multistatus>

Visualization

Proposal for visualization in CMS Dashboard based on ATLAS implementation



We are currently also looking in Elasticsearch+Kibana based implementation

Fermilab

Potential areas of collaboration with WLCG/ATLAS

CMS specific	WLCG/ATLAS Common
TFC (site configuration)	Storage technologies
Data storage namespace	Storage dump formats
Authenticated upload	Middleware software infrastructure
Monitoring configuration	Visualization infrastructure

- CMS SpaceMon will clearly benefit from WLCG common infrastructure and tools for storage information providers and visualization
- CMS specifics tasks, such as:
  - translating local storage areas to a global logical namespace
  - defining and maintaining aggregation parameters
  - site specific authentication and roles
  - monitoring configuration

  need to be done on the experiment side.

Deployment campaign

Issues encountered during this first phase of deployment can be categorized into three groups:

1. Questions from sites about why they need to provide storage usage information and at what level of detail
2. Authentication problems uploading the information to the central data service
3. The long time it takes to take a dump for some storage systems.

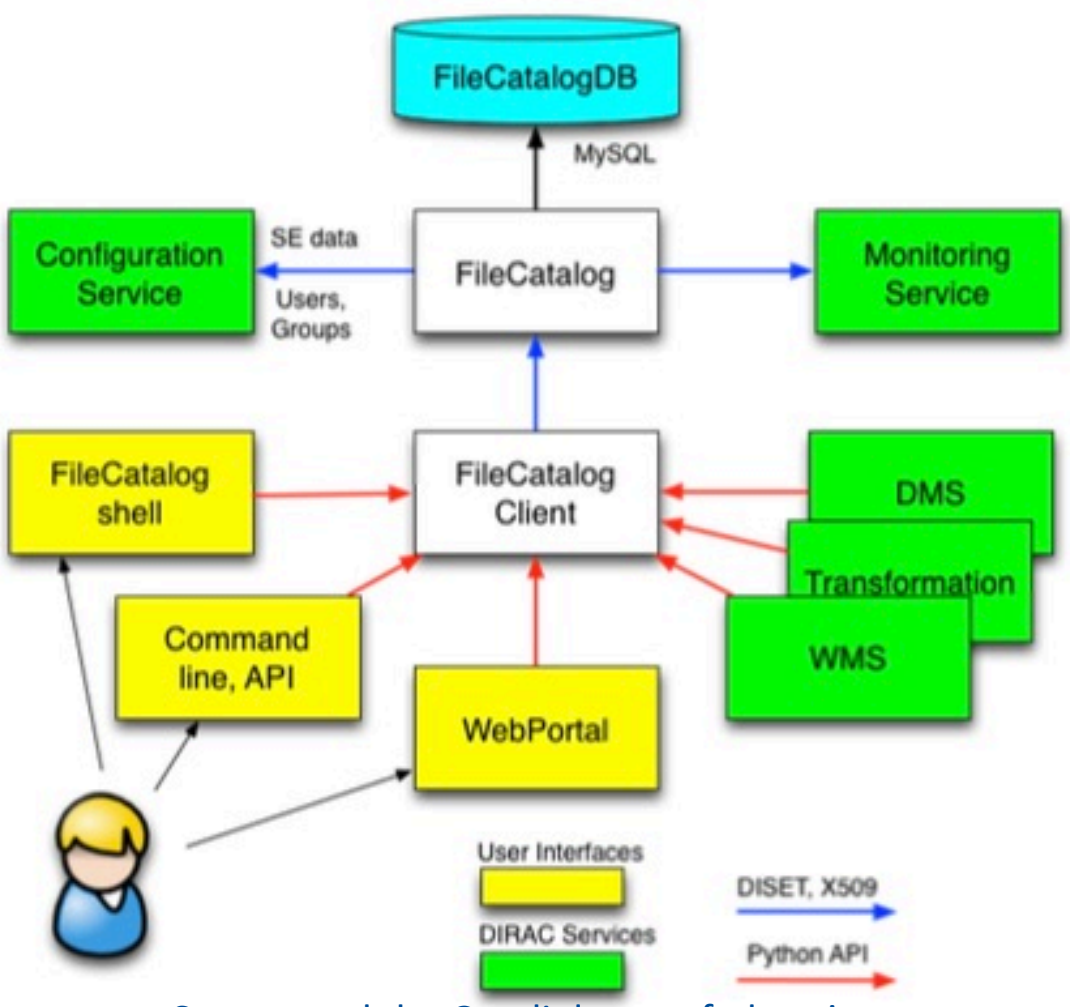
Also some privacy and security concerns were raised by the sites.

Summary

- In order to effectively organize storage and processing of the data, the LHC experiments require a reliable and complete overview of:
  - the storage capacity in terms of the occupied and free space
  - the storage shares allocated to different computing activities
  - possibility to detect “dark” data that occupies space while being unknown to the experiment’s file catalog.
- CMS developed Space Monitoring system based on the storage dumps using formats, recommended by WLCG
- We are currently looking for areas of common interest and further collaborative effort within WLCG Experiment support team, including CMS, ATLAS and potentially LHCb.

Backup slides

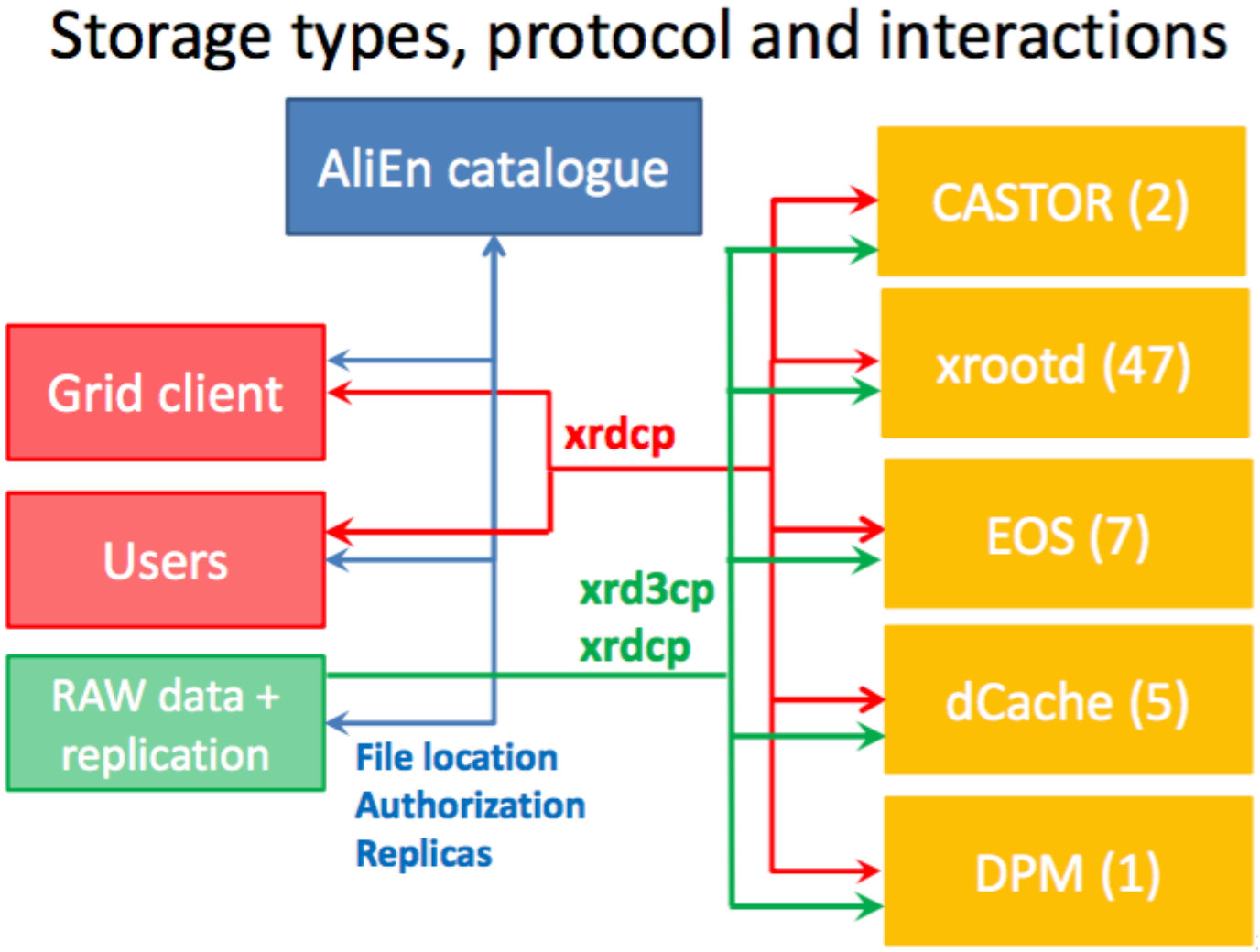
The LHCb DIRAC File Catalog



The DIRAC Data Management System and the Gaudi dataset federation
http://dx.doi.org/10.1088/1742-6596/664/4/042025

Fermilab

Alice data management



Summary of the Experiments Data Management Inputs DM WLCG Workshop 2016 Lisbon (S.Campana)